

# 웹사이트 성능 향상을 위한 웹페이지 분석 도구

김현표, 이도헌  
전남대학교 전산통계학과  
e-mail:{hpkim, dhlee}@dmlab.chonnam.ac.kr

## A Web-Page Analysis Tool to Improve Web-Site Performance

Hyun-Pyo Kim, Doheon Lee  
Dept of Computer Science and Statistics,  
Chonnam National University

### 요 약

웹에 등장한 다양한 웹사이트는 사용자의 정보획득 욕구를 충족시켜주는 최고의 정보 자원이 되었다. 하지만 최근 웹사용자의 경향이 보다 더 좋은 성능에서 정보를 얻고자 하여 정보 전달 능력이 낮은 사이트들은 양질의 정보를 보유하고도 사용자들에게 외면 당하고 있는 추세이다. 본 논문에서는 웹사이트의 성능을 향상시키기 위해 주방문자의 정보와 전송 시간에 관련된 여러 가지 요소, 웹페이지 구성요소의 중요도, 전송 속도와 사용자 만족도 간의 관계 등을 이용하여 사이트의 정보 전달 능력을 개선시킬 수 있는 웹페이지 분석 도구를 제안하고자 한다.

### 1. 서론

웹의 발달로 인하여 인터넷상에는 사용자들의 정보 획득 욕구를 충족시켜줄 수 있는 다양한 웹사이트가 등장하였다. 그러나 최근의 웹사용자들은 단순히 정보를 얻는 것이 아니라 더 좋은 성능으로 정보를 획득하기를 요구하고 있다. 이를 위해 웹사이트 관리자들은 다양한 방법으로 성능 개선을 시도하고 있다.

본 논문에서 제안하는 웹페이지 분석 도구는 웹 문서의 구조 개선을 통해 웹사이트의 성능을 향상시키기 위한 목적으로 설계되었다. 분석 과정은 제작된 웹페이지의 유효 파일크기[1]와 방문자 통계, 웹사용자 만족도, 웹페이지 구성요소의 중요도 등의

정보를 이용하여 웹페이지의 구조를 개선한 후 전송 속도를 향상시켜 사용자에게 빠른 시간에 정보를 전달하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 웹성능 향상을 위한 방법과 웹페이지의 유효 파일크기에 대해 알아보고 웹사이트 방문자 통계를 구하기 위해 수행되어지고 있는 웹 로그파일 분석에 대해 알아본다. 3장에서는 본 논문에서 웹성능의 주요 평가 기준으로 사용하고 있는 웹페이지 전달효과 시간, 웹페이지 사용자 만족도, 웹페이지 구성요소의 중요도 등의 개념을 논하고 위 세 가지 기준을 이용한 분석도구를 제안하며 마지막으로 4장에서는 분석도구의 이용 방향과 향후 연구 방향의 제시로 결론을 맺고자 한다.

## 2. 관련 연구

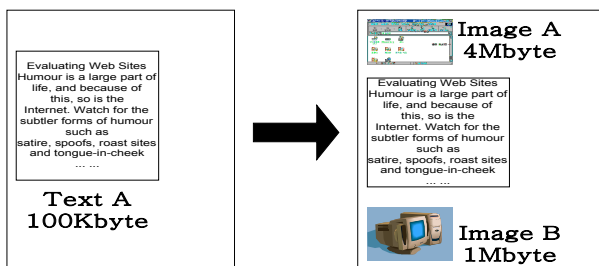
### 2.1 웹 성능 향상

웹에 대한 성능 평가는 웹페이지가 포함하고 있는 정보의 질적 수준을 측정하는 논리적 평가와[2] 전송시간에 근거한 속도를 측정하는 물리적 평가로 [3] 크게 나뉘어진다. 정보의 질적 평가는 문서의 내용에 관한 제작자의 관련 지식 수준, 내용의 신뢰성, 정확성 등 웹페이지가 전달하고자 하는 정보에 대한 충실도가 주요 평가 기준이 되며 전송속도 측면에서의 평가는 동일한 정보의 양을 빠른 시간 내에 사용자에게 전달하기 위해 사용된 통신망의 전송 능력이 주요 평가 기준으로 작용한다.

웹성능 향상은 통신망의 성능 향상과 웹문서의 구조 변경을 이용한 성능 향상[4]이라는 두 가지 방법으로 연구되어지고 있다. 통신망의 개선을 이용한 성능향상 방법은 성능 측정의 주 요소가 되는 지연시간(latency)과 대역폭(bandwidth)을 고려하여 프로토콜의 재설계나 통신 설비의 교체를 통해 성능의 향상을 시도하고 웹문서의 구조개선 방법에서는 웹페이지의 구성 요소 중 많은 전송시간을 필요로 하는 요소의 크기를 축소하거나 다른 요소로 대체하는 방법을 이용하여 성능 향상을 가능하게 한다.

### 2.2 웹페이지의 유효 파일크기

웹페이지의 유효 크기는 전송에 영향을 주는 구성요소들의 파일크기를 더한 값과 같다.



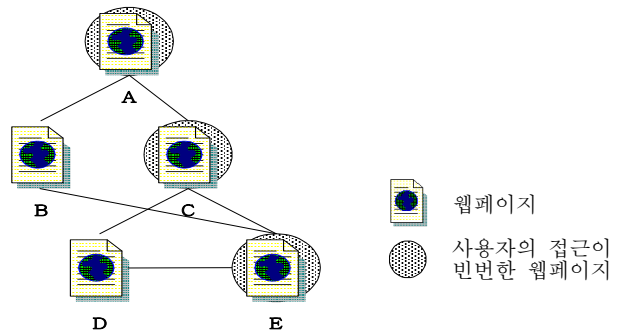
[그림 1] 구성 요소 추가에 의한 유효 파일크기 변화

[그림 1]은 최초 텍스트만으로 제작된 100Kbyte 크기의 HTML 문서에 두 개의 이미지 파일이 추가되어 확장한 것이다. 이 웹페이지의 파일크기는 이미지 파일을 추가하기 사용된 HTML 태그문자로 인해 약 50~80byte정도 증가하였으나 전송에 관련된 유효 파일크기는 최초 100Kbyte에서 4Mbyte와

1Mbyte의 그림 파일이 추가되어 기존의 HTML 문서의 크기 100Kbyte에 그림 파일들의 크기가 합하여진 5.1Mbyte로 증가한다.

### 2.3 웹 로그의 분석을 이용한 웹사이트 개선

웹 로그 파일은 웹서버에 접속한 클라이언트의 IP 주소, 접근 시간, 접근한 페이지의 URL 등의 정보를 포함하고 있다. 이 로그 파일의 분석 결과를 이용하여 웹사이트 내에서 가장 빈번히 접근되는 페이지나 사용자의 이동 패턴 등을 파악[5]할 수 있으므로 웹사이트의 정보를 효과적으로 전달하기 위해 로그 파일을 이용하고 있다.



[그림 2] 웹사이트의 일반적인 구조

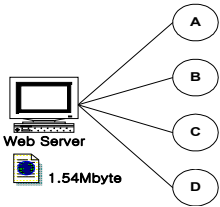
[그림 2]는 일반적인 웹사이트의 구조이다. 이 웹사이트의 로그파일을 분석한 결과 방문자의 접근이 웹페이지 A, C, E 순서로 이동하는 성향이 가장 많았다고 하자. 사이트 관리자는 방문자의 관심도가 높은 페이지에 다른 페이지보다 더 많은 정보를 포함시키거나, 사이트를 재구성하여 A페이지에 E페이지로 직접 이동할 수 있는 링크를 두어 연결 할 수 있다. 또한 A, C, E의 페이지의 구성을 변경하여 접근자에게 웹사이트의 정보를 빠르게 전달하는 방법 등으로 웹사이트의 성능을 향상시킬 수 있다.

## 3. 웹 성능평가의 기준과 분석 도구 제안

웹사이트의 성능은 다양한 방법으로 분석할 수 있으나 본 논문에서는 웹페이지의 전달효과 시간, 전송속도에 따른 사용자 만족도, 웹페이지 구성요소의 중요도를 이용하여 분석하고자 한다. 또한 제안하는 분석도구는 웹사이트가 제공하는 웹페이지의 유효 파일의 크기를 측정하여 사용자의 만족도와 비교한 후 중요도가 떨어지는 구성 요소들의 제거와

대체를 통해 웹 성능을 향상시키는 것을 목적으로 한다.

### 3.1 웹페이지 전달효과 시간



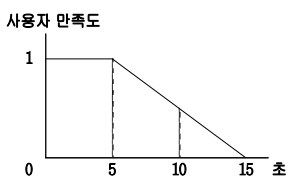
[그림 5] 웹서버 방문자 집단

[표 1] 웹서버 방문자 집단 정보

접근 집단	집단 분포	전송 속도	전송 시간
A	40%	1.54Mbps	1초
B	30%	128Kbps	12초
C	20%	64Kbps	24초
D	10%	33Kbps	48초

[그림 3]은 하나의 웹사이트에 접근하는 방문자들을 접속 지역과 사용하는 통신장비의 대역폭별 등 다양한 기준으로 그룹화하여 집단 A, B, C, D로 나타내었다. [표 1]은 [그림 3]에서 그룹화된 방문자 집단을 전체 사용자중에서 차지하는 분포, 사용하는 통신 장비에 따른 이론적인 전송속도와 웹서버에 있는 1.54Mbyte 크기의 웹페이지를 전송 받는데 걸리는 전송 시간 등의 정보를 분석한 결과이다. 각 집단이 웹 서버의 정보를 전송 받는 시간은 접근 집단이 사용하는 통신망의 대역폭에 따라 크게 다르나 웹사용자의 집단 분포와 평균 전송 시간을 이용하여 가중치가 부여된 전송시간을 계산하면 1.54Mbyte의 유효 파일크기를 갖는 웹페이지는 약 13.6초(40%\*1초+30%\*12초+20%\*24초+10%\*48)의 시간에 전체 사용자에게 전달된다. 본 논문에서는 위와 같이 사용자의 통신망 대역폭에 의한 전송시간과 접근 집단의 분포를 이용하여 구해진 시간을 웹페이지의 전달효과 시간이라 한다. 또한 동일한 방법으로 전송 속도와 집단의 분포를 이용하여 전달효과 대역폭을 구할 수 있다. 위의 예에서 전달효과 대역폭은 670.5Kbps(40%\*1540+30%\*128+20%\*64+10%\*33)이다.

### 3.2 웹페이지 접근자 만족도



[그림 6] 웹페이지 전송시간-접근자 만족도 그래프

[표 2] 웹페이지 전송시간과 접근자 만족도간의 관계식

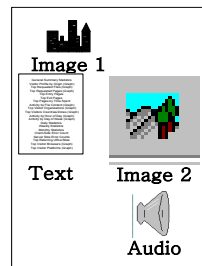
1	, 0 < X ≤ 5
$Y = (15-X)/10$	, 5 < X ≤ 15
0	, 15 < X ≤ ∞
X : 웹 페이지의 전송 시간	
Y : 사용자 만족도	

공학심리학적으로 웹사용자는 웹페이지에 접근을

시도한 후 약 5초후부터 지루함을 느끼기 시작하고 전송 속도가 15초 이상 경과할 경우 다른 사이트로의 이동을 시도한다. [그림 4]는 웹페이지 전송시간에 대해 접근자가 느끼는 만족도에 관한 그래프이다. [그림 4]의 그래프는 전송시간이 5초 이상 15초 이하의 시간까지 접근자의 만족도가 단조 감소하는 것으로 나타내어진 것이고 본 논문에서는 [표 2]와 같이 전송시간과 만족도간의 관계를 1차 방정식으로 표현하였다.

### 3.3 웹페이지 구성요소의 중요도

웹페이지 구성요소의 중요도는 웹페이지의 정보 전달과정에서 어느 정도 중요하게 사용되는가를 수치 값으로 표현한 것이다.



[그림 7] 네개의 요소로 구성된 웹페이지

[표 3] 웹페이지 구성요소간의 전송 시간 정보와 사용자 부여 우선 순위 정보

	Image1	Text	Image2	Audio
크기	2KB	1KB	2MB	40M
우선 순위	1	2	3	4
중요도	4/10	3/10	2/10	1/10
전송예상 시간(초)	0.002	0.001	2	40

전송 속도 1Mbps 기준

[그림 5]는 두 개의 이미지 파일, 텍스트 영역, 한 개의 오디오 파일로 제작된 웹페이지로써 각각의 요소는 웹페이지의 정보를 전달하기 위해 필요한 순서에 따라 우선 순위를 부여받는다. 이 우선 순위를 이용하여 전체 합이 1이 될 수 있도록 구성요소의 중요도를 나타낼 수 있다. 요소별 중요도는 웹페이지의 구성 요소가 m개이고 우선 순위가 n인 경우  $(m+1-n)/(1+2+\dots+m)$ 의 값으로 부여할 수 있다. [표 4]는 [그림 5]에서 나타내는 웹페이지에 구성요소들의 파일크기 정보와 각 요소에 우선 순위 부여하여 중요도를 구한 결과이다. [표 3]에 의하면 [그림 5]에서 나타내는 웹페이지의 전송속도는 42.003초기 소요된다.

[그림 5]의 웹페이지에서 성능 개선을 위해 가장 우선 순위 낮은 오디오 파일을 동일 정보를 갖는 이미지파일이나 텍스트로 대체할 경우 웹페이지의 전달효과는 다소 감소하나 전송속도 측면에서 큰 이득을 얻을 수 있다. 예를 들어 오디오 파일을 4Mbyte의 이미지 파일로 대체할 경우 전달효과는 1/2로 줄

고 하고 1Kbyte 텍스트 파일로 대체할 경우 전달효과는 1/4로 줄어든다고 가정하자. 이미지 파일로 대체한 경우 중요도는 0.95로 0.5만큼 감소하고 텍스트로 대체한 경우 중요도는 0.925로 0.75만큼 감소한다. 하지만 전송 시간의 측면에서는 이미지 파일로 대체 시에는 6.003초 텍스트로 대체 시에는 2.003초의 전송 시간으로 최초 42.003초와는 큰 차이를 보인다.

### 3.4 분석 도구 제안

본 논문에서 제안하고자 하는 분석도구는 최초 웹사이트의 주방문자와 타겟으로하는 방문자 정보를 입력받은 후 집단화하고 웹사이트의 전달효과 시간을 구하는 과정을 수행한다. 이 전달효과 시간을 사용자 만족도와 비교하여 웹사이트의 적절한 유효 파일 크기를 구하고 만약 제작한 웹사이트의 유효 파일크기가 적정값 이상일 경우 웹사이트 구성요소의 중요도를 이용하여 구성요소를 대체하는 과정을 수행하고 각 과정의 상세 설명은 다음과 같다.

#### 1. 전달효과 대역폭 계산

웹사이트의 주된 방문자와 웹사이트가 타겟으로 하는 방문자의 정보를 입력하여 방문자의 지역, 사용하는 통신 장비 등의 정보로 집단화하고 집단의 정보를 이용하여 전달효과 시간과 대역폭을 구한다.

#### 2. 전달효과 대역폭과 사용자 만족도 비교

전달효과 대역폭과 [그림 4]의 사용자 만족도를 이용하여 최대 유효 파일크기를 구한다.

#### 3. 중요도에 의한 구성요소 변경 및 삭제

제작된 웹사이트의 크기가 최대 유효 파일크기를 초과했을 경우 구성요소 중 중요도가 낮은 순서의 요소부터 대체하거나 삭제 시에 변경될 중요도와 전달효과 시간을 제시하고 복수의 요소가 같은 중요도를 가질 경우 유효 크기가 큰 요소부터 수행한다.

### 4. 결론 및 향후 연구 과제

현재 구축되고 있는 웹사이트들은 다양한 기술과 사용자 분석을 통해 많은 사용자가 접근하기를 유도하고 있다[6]. 이를 위해 웹 로그파일 분석이나 유효 파일크기를 축소하는 등의 다양한 방법들을 이용하여 웹사이트의 성능 향상을 시도하고 있다.

본 논문에서 제안한 웹사이트 분석도구는 웹사이

트 구축이나 재설계에 사용자로부터 입력받은 방문자와 타겟으로하는 방문 예상자의 정보와 웹페이지를 분석하여 얻은 웹페이지의 전달 속도와 해당 페이지가 포함하고 있는 구성요소들의 중요도에 따라 웹페이지의 구조를 변경하여 웹사이트가 포함하고 있는 정보가 웹 사용자에게 전달될 때까지 소요될 전송 시간을 단축함으로써 웹성능을 개선하고자 한다.

향후 효과적인 웹사이트의 전달을 통한 성능 향상을 위해 본 논문에서 제안한 웹사이트 분석도구를 구현하고자 한다.

### 5. 참고 문헌

- [1]. Joe Touch, John Heidemann, and Katia Obraczka. "Analysis of HTTP Performance", Research Report 98-463, USC/Information Sciences Institute, August, 1998.
- [2]. Craig Braham, "A Student's Guide to Research with the WWW", "<http://www.slu.edu/department/english/research>", March, 1997.
- [3]. Robert Harris, "Evaluating Internet Research Sources", "[http://www.sccu.edu/faculty/R\\_Harris/evalu8it.htm](http://www.sccu.edu/faculty/R_Harris/evalu8it.htm)", November, 1997.
- [4]. 김용운, "Performance Improvements in WWW", the 5th WWW Work-shop 1997, "<http://pec.etri.re.kr/~qkim/HTTP/performance.html>"
- [5]. 남도원, 이동하, 서동렬, 이진영, "웹 로그에서의 사용자 접근 패턴 분석", 정보과학회 인간과 컴퓨터 상호작용 연구회 회보 7권 1호, HCI'98 학술대회 발표논문집, 1998년 2월. pp. 160-165.
- [6]. Internet Paradigm Communications, "Understanding Web Performance", "<http://www.interdig.com/hosting/features/performance.html>"